Lecture 13, MTH 400, Fall 2024

Storytelling with Data

Visual perception plays a crucial role in the effectiveness and interpretation of data visualizations. Our visual system has certain inherent characteristics and limitations that can impact how we perceive and interpret visual information. Understanding these factors helps in designing effective data visualizations that accurately convey the intended message. Here are some key aspects of visual perception that affect data visualizations:

1. Pre-attentive Processing: Our visual system automatically processes certain visual attributes such as color, size, and shape without conscious effort. These pre-attentive features attract our attention and allow for quick perception and identification. Utilizing these features strategically in data visualizations helps highlight patterns or outliers.
2. Gestalt Principles: The Gestalt principles describe how we perceive and organize visual elements. These principles include proximity, similarity, continuity, closure, and figure-ground relationships. Applying these principles in data visualizations aids in grouping and structuring visual elements to facilitate interpretation and understanding.
3. Visual Hierarchy: Visual hierarchy refers to the arrangement of visual elements in terms of their importance and significance. By utilizing techniques such as size, color, or position, we can establish a visual hierarchy that guides the viewer's attention and emphasizes key information in the visualization.
4. Color Perception: Color plays a crucial role in data visualizations. It can be used to encode different categories, highlight important information, or represent quantitative values. However, color perception can be influenced by factors such as color blindness or cultural associations, so it is important to select appropriate color palettes and consider accessibility requirements.
5. Visual Encoding: The choice of visual encodings, such as position, length, angle, and area, can significantly impact how we interpret and compare values in a visualization. For example, using a length encoding for a quantitative variable is more accurate than using an angle encoding.
6. Cognitive Load: Visualizations that are overly complex or cluttered can overwhelm the viewer and hinder understanding. Minimizing cognitive load by simplifying the visualization and presenting information in a clear and concise manner helps facilitate comprehension.
7. Context and Reference: Providing appropriate context and reference points in data visualizations helps viewers understand the scale, scope, and relevance of the presented information. Labels, legends, titles, and annotations contribute to the overall interpretability of the visualization.
8. Visual Attention and Focus: People have limited attention spans and tend to focus on specific areas of a visualization. Understanding how visual attention operates helps in designing visualizations that guide viewers to the most important or relevant aspects of the data.
9. Cultural and Contextual Factors: Visual perception can be influenced by cultural backgrounds, prior experiences, and individual preferences. Consideration of the target audience and their cultural and contextual factors helps ensure that the visualization is effective and meaningful.

By understanding these aspects of visual perception, data visualization designers can create visual representations that are more intuitive, engaging, and effective in conveying the intended information and insights.

Telling stories with data is an effective way to communicate insights, trends, and narratives hidden within the data. It goes beyond presenting raw numbers and charts to create a compelling narrative that engages and resonates with the audience. Here are some key strategies to tell stories with data:

1. Identify the Audience: Understand your target audience, their background, knowledge level, and interests. This knowledge helps shape the narrative and select appropriate visuals and storytelling techniques that resonate with the audience.
2. Define a Clear Narrative: Start by defining the main message or story you want to convey through the data. Craft a clear and concise narrative that has a logical flow and builds towards a central theme or takeaway.
3. Use Context and Background: Provide context and background information to set the stage for the data story. Explain the problem, the data sources, and any relevant background information that helps the audience understand the significance of the data.
4. Engage with Anecdotes and Examples: Incorporate anecdotes, examples, or real-life stories that connect with the data and make it relatable. Personal stories and case studies can bring the data to life and make it more meaningful for the audience.
5. Structure the Data Story: Organize the data story into sections or chapters that lead the audience through a narrative arc. Build tension, introduce key points, and provide resolutions or conclusions to maintain engagement and interest.
6. Visualize the Data: Select appropriate visualizations that support and enhance the narrative. Use charts, graphs, infographics, and other visual elements to present the data in a visually appealing and easily understandable manner. Ensure that the visualizations align with the story and effectively convey the intended message.
7. Simplify Complex Concepts: Break down complex concepts or technical jargon into simpler terms that are accessible to the audience. Use clear and concise language to explain the data and avoid overwhelming the audience with unnecessary details.
8. Create Compelling Headlines and Titles: Craft compelling headlines or titles that capture the essence of the data story and grab the audience's attention. Headlines should be concise, intriguing, and provide a glimpse of the insights or narratives to come.
9. Utilize Storytelling Techniques: Incorporate storytelling techniques such as introducing characters (e.g., personas), creating a plot (e.g., problem-solution), or using metaphors and analogies to help the audience connect with the data on an emotional level.
10. Provide Actionable Insights: Conclude the data story with actionable insights or recommendations. Help the audience understand the implications of the data and provide guidance on how to apply the insights to real-world situations.

Remember, telling stories with data is not just about presenting the data itself but weaving a narrative that engages, educates, and inspires the audience. By combining data, visuals, storytelling techniques, and audience-centric communication, you can effectively convey the significance and impact of the data in a compelling and memorable way.

Q. What would be considered examples of advanced data visualizations?

A. Advanced data visualizations go beyond simple bar charts or scatter plots to represent complex and multivariate data in innovative and informative ways. These visualizations often require sophisticated techniques and specialized tools to create. Here are a few examples of advanced data visualizations:

1. Sankey Diagrams: Sankey diagrams are used to depict flows and relationships between different entities. They are particularly effective in visualizing processes, energy flows, or network connections. The width of the flows represents the quantity or magnitude of the data.
2. Chord Diagrams: Chord diagrams illustrate the relationships and connections between entities in a circular layout. They are commonly used to visualize relationships in network analysis, such as interconnections between different nodes or groups.

3. Parallel Coordinates: Parallel coordinates plots display multivariate data by representing each variable as a vertical axis and connecting the data points with lines. This visualization allows for the exploration of patterns, correlations, and outliers across multiple dimensions simultaneously.
4. Treemaps: Treemaps use nested rectangles to represent hierarchical or tree-structured data. Each rectangle's size represents a quantitative value, and the color or shading can be used to encode additional information. Treemaps are useful for visualizing hierarchical data, such as file systems, organizational structures, or market segmentation.
5. Heatmaps and Dendrograms: Heatmaps combine a matrix of colors to represent the values of a two-dimensional dataset. They are often used to display patterns, correlations, or clustering in data. Dendrograms are frequently paired with heatmaps to show hierarchical clustering in the rows or columns of the heatmap.
6. Force-Directed Graphs: Force-directed graphs use physics-based simulations to visualize relationships between nodes. The nodes are connected by edges, and the layout positions are determined by attractive and repulsive forces. Force-directed graphs are useful for visualizing social networks, collaborations, or any network of interconnected entities.
7. Geographic Information Systems (GIS): GIS visualizations combine geospatial data with other variables to create interactive maps. They are commonly used to explore and analyze spatial patterns, such as population density, disease outbreaks, or environmental factors.
8. Interactive Dashboards: Interactive dashboards provide a user-friendly interface to explore and interact with complex datasets. They often include a combination of visualizations, filters, and controls that allow users to drill down into the data and gain insights dynamically.
9. 3D Visualizations: Three-dimensional visualizations add depth to data representation, allowing for the exploration of spatial relationships or volumetric data. They are commonly used in fields such as medical imaging, geology, or computer graphics.
10. Augmented Reality (AR) and Virtual Reality (VR): AR and VR technologies offer immersive data visualizations that enable users to interact with data in a three-dimensional or virtual environment. They are particularly useful for simulations, training, or complex data exploration.

These are just a few examples of advanced data visualizations. The choice of visualization technique depends on the characteristics of the data, the research questions, and the desired insights. Advanced data visualizations often require specialized software tools or programming libraries to create, and they can provide unique perspectives and deeper understanding of complex datasets.

Creating effective and clear graphs is crucial for accurately communicating data insights. However, there are several common pitfalls to avoid to ensure your visualizations are both informative and easy to understand. Here are some things to avoid when making graphs:

1. Overloading with Information
- Too Much Data: Avoid including too many data points or series, which can make the graph cluttered and difficult to read.
- Excessive Annotations: Overloading the graph with too many labels, lines, or annotations can distract from the main message.
2. Poor Choice of Graph Type
- Inappropriate Graph Types: Choose the right type of graph for your data. For example, using a pie chart for large numbers of categories can be confusing.
- 3D Graphs: Avoid using 3D charts as they can distort the data and make it hard to read. Stick to 2D charts for clarity.
3. Misleading Scales

- Non-Zero Baselines: Starting the y-axis at a value other than zero can exaggerate differences and mislead the viewer.
- Inconsistent Scales: Ensure that all axes are consistently scaled. Using different scales on a multi-series graph can distort the comparisons.

4. Poor Design Choices

- Overuse of Colors: Using too many colors can be distracting and make it difficult to distinguish between data series. Stick to a limited and meaningful color palette.
- Inconsistent Colors: Ensure that colors are used consistently across multiple graphs to represent the same categories or series.
- Poor Contrast: Ensure there is sufficient contrast between the text, data points, and background to make the graph readable.

5. Lack of Context

- Missing Labels: Always label your axes and provide a title for the graph to explain what the data represents.
- No Legend: If you have multiple data series, include a legend to explain what each series represents.

6. Ignoring Accessibility

- Color Blindness: Avoid using color combinations that are difficult to distinguish for color-blind users. Use tools like ColorBrewer to choose accessible color schemes.
- Font Size: Ensure that text is large enough to be readable on all devices and in print.

7. Data Distortion

- Inaccurate Data Representation: Ensure that the data is represented accurately without any distortion or manipulation.
- Cherry-Picking Data: Avoid selectively presenting data that supports a particular narrative while ignoring data that provides a fuller picture.

8. Ignoring Best Practices

- Gridlines: Use gridlines sparingly to avoid clutter. Light, subtle gridlines can help with readability without overwhelming the graph.
- Aspect Ratio: Maintain an appropriate aspect ratio to avoid distorting the data. The default settings in many software packages are often a good starting point.

Examples of Good Practices
To illustrate these points, here are some examples of good practices in graph making:

Clear and Simple Line Chart

```
library(ggplot2)
data <- data.frame(
  year = c(2010, 2011, 2012, 2013, 2014, 2015),
  value = c(100, 150, 200, 250, 300, 350)
)

ggplot(data, aes(x = year, y = value)) +
  geom_line(color = "blue") +
  geom_point() +
  labs(title = "Annual Growth", x = "Year", y = "Value") +
  theme_minimal()
```

Avoiding Clutter in Bar Charts

```
data <- data.frame(
  category = c("A", "B", "C", "D"),
  value = c(3, 12, 5, 18)
)

ggplot(data, aes(x = category, y = value)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Category Values", x = "Category", y = "Value") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

By avoiding these common pitfalls and adhering to best practices, you can create clear, accurate, and effective graphs that effectively communicate your data insights. Always strive for simplicity, accuracy, and clarity in your visualizations.

Resources:

1. https://www.youtube.com/channel/UCjhGIILWNloXJdR2NTCBMlA
2. https://ourcodingclub.github.io/tutorials/dataviz-storytelling/
3. https://www.storytellingwithdata.com/
4. https://github.com/Saurav6789/Books-/blob/master/storytelling-with-data-cole-nussbaumer-knaflic.pdf
5. https://www.oreilly.com/library/view/storytelling-with-data/9781119002253/f_07.xhtml